Maintaining Mappings Valid between Dynamic KOS

Julio Cesar dos Reis

Maintaining Mappings Valid between Dynamic KOS

Julio Cesar Dos Reis
CR SANTEC, CRP Henri Tudor and LRI, University of Paris-Sud XI
6, Av. des Hauts-Fourneaux, L-4362 Esch-sur-Alzette, Luxembourg
julio.dosreis@tudor.lu

Abstract. **K**nowledge **O**rganization **S**ystems (KOS) and the existing mappings between them have become extremely relevant in semantic-enabled systems especially for interoperability reasons. KOS may have a dynamic nature since knowledge in a lot of domains evolves fast, and thus KOS evolution can potentially impact mappings, turning them unreliable. A still open research problem is how to adapt mappings in the course of KOS evolution without re-computing semantic correspondences between elements of the involved KOS. This PhD study tackles this issue proposing an approach for adapting mappings according to KOS changes. A framework is conceptualized with a mechanism to support the maintenance of mappings over time, keeping them valid. This proposal will decrease the efforts to maintain mappings up-to-date.

Keywords: mapping evolution; mapping adaptation; mapping maintenance.

1 Introduction

Knowledge **O**rganization **S**ystems (KOS) aim at encompassing all types of conceptual models for organizing knowledge [1] as, for example, semantic networks, ontologies, taxonomies and thesauri. In various contexts and domains, such as the Semantic Web (SemWeb) and Bioinformatics, it is necessary to have a combined use of different KOS, since a unique KOS is not able to cover the totality of a domain due to its size and complexity. Mappings representing semantic correspondences between elements belonging to different KOS therefore need to be established.

The highly dynamic aspect of the knowledge leads to frequent KOS changes. Klein [3] proposed a first categorization of changes, which can affect ontologies, dividing them into atomic and complex operations. The first refers to the change of only a single specific element (*e.g.*, concepts, attributes and relationships) while the second denotes operations that are composed of multiple atomic ones. The impact of these changes on mappings associated to KOS has not been deeply studied. Actually, KOS evolution challenges the reliability of dependent artifacts such as mappings, in the sense that changes affecting KOS elements may invalidate existing mappings. This requires mappings to be adequately maintained over time. Nevertheless, how to adapt mappings impacted by KOS evolution as automatic as possible, without re-computing the whole set of mappings each time a KOS evolves, is still an open research problem. Many research questions arise in the context of this problem: (1) How to perform mapping adaptation taking the way KOS evolve into account? (2) What information regarding mappings and KOS evolution is necessary to support the mapping adaptation? (3) How to correlate different types of KOS changes with actions suited to adapt mappings? (4) How might the different types of semantic relations of mappings be taken into consideration?

Maintaining mappings valid over time is crucial since various applications may rely on them [2]. In the SemWeb context, for instance, up-to-date mappings could allow more trustable semantic searches over integrated ontologies in the Web [4]. In other domains, such as the biomedicine, mappings are very important to support data integration among different applications [5]. Usually, hundreds of thousands of mappings are explored by applications such as the Unified Medical Language System (UMLS). Therefore, after releasing new KOS versions, re-computing the whole set of mappings is a time-consuming task demanding huge efforts of validation.

The aim through this PhD study is to define a framework coping with the mapping maintenance problem between dynamic KOS. The proposed approach developed in the framework is to adapt mappings relying on the exploitation of information derived from KOS evolution, combined with information coming from existing mappings. We aim at considering different types of semantic relations ($=, \leq, \geq, \approx$) in mappings.

The remainder of this article is organized as follows: Section 2 presents the state-of-the-art. Section 3 describes the proposed approach including the research methodology and the evaluation method. Section 4 presents the results achieved so far and the future work envisaged.

2 The state of the art

Although significant research efforts in the past years have coped with issues related to ontology evolution [6], the understanding of the impact of this evolution in dependent artifacts such as mappings has received very little attention. We organize the different approaches coping with the maintenance of mappings in two main categories. The first category tackles the problem by re-calculating mappings. The most naïve approach is the full re-calculation of the set of mappings, which does not consider any information from KOS changes or mappings. Nowadays, there is a high frequency of new KOS versions, and usually the rate of KOS evolution does not justify a full re-calculation [7]. A partial re-calculation approach was proposed by Khattak *et al.* [8] re-creating only those mappings associated to ontology elements which had changed. Matching algorithms are used to perform a new alignment between those changed elements and the whole target ontology. However, the size of KOS still challenges the compromise between precision and recall of available techniques for mapping calculation [2]. Partial re-calculation slightly minimizes the efforts of validation.

The second category concerns approaches attempting to adapt mappings after KOS evolution. KOS changes are usually used to support adaptation of mappings without performing re-calculation. The first propositions appeared in the context of database schema mappings [9]. For ontologies, Martins & Silva [10] propose that evolution of mappings should behave similarly with the strategies applied for ontology evolution.

More recently, aiming at better understanding mapping evolution, Groß *et al.* [11] empirically investigated the evolution of life sciences ontology mappings. In fact, it is still unknown how to fully perform mapping adaptation as automatic as possible according to KOS evolution. The influence of KOS changes on how mappings should change deserves deeper investigations and various research problems remain open. For instance, considering the change in the semantic relation type of a mapping as a possible mapping adaptation action is still an issue. It is also crucial to conduct further investigations to better understand the impact of complex changes operations (*e.g.,* split and merge of concepts) on mappings for their adequate adaptation.

3 Approach and Methodology

This research relies on the hypothesis that there is a correlation between changes affecting KOS elements and the evolution of their associated mappings, which has been observed in experimental studies. In this sense KOS evolution shall be well described for supporting the adaptation of mappings. This is the characterization of a refined categorization of underlined KOS (complex) change operations (the most fine-grained types of KOS changes) containing information judged important for the adaptation of mappings. We determine that as *Change Patterns* (CPs) in a way to recognize different behaviours of changes between KOS versions and a richer context to adapt mappings. Different types of split complex operations are examples of CPs. These are expressed as distinct types of atomic change operations (*i.e.*, addition and removal of KOS elements) as well as KOS complex change, including whenever possible, information regarding the semantic and structural impact of these changes.

We have identified different behaviours of complex changes such as split and merge of concepts. These behaviours are recognized according to a categorization of semantic similarity shared between concepts in a change operation. We also consider how involved concepts in the change are structurally organized. For instance, whether merged concepts were related through an 'is_a' relationship or whether they were sibling concepts. These aspects are further explored for the mapping adaptation.

The proposition is to adapt mapping elements such as the source element of the mapping, and/or the type of its semantic relation supported by information from the CPs that have affected the mapping combined to information coming from mappings. *Mapping Adaptation Actions* are proposed representing different strategies of mapping adaptation to change the adequate mapping elements, for instance, to adapt mappings associated to a removed concept transferring them to parent or sibling concepts (two different actions). In order to know the most appropriate action to be taken for each mapping independently, CPs information and identified elements used to establish the mapping including its semantic status are taken into account. These must represent the necessary conditions to model in which situation an adaptation action shall be applied. *Heuristics* in the proposal accounts for the modeling and the formalization of these conditions, thus expressing the correlations between information from mappings and CPs with the adaptation actions.

As an example, if a deletion of an attribute affects a concept and this attribute was identified as crucial for establishing an associated mapping of the concept concerned, then such mapping is removed. Also, if a complex change like a split of concept was identified resulting in new sub concepts (an example of CP), and an early mapping with the relation of less general type (≤) was associated to the old unsplit concept, then the sub-concepts may inherit this mapping, keeping the same semantic type. Note that how the mapping is adapted is dependent on the structural organization of the concepts in the change combined with information from the mapping.

The research methodology conducted firstly observe empirically the evolution of various KOS from the biomedical domain, and the way different types of KOS (complex) change have impacted the behavior of existing official mappings. The proposed approach in the framework is grounded on the results of these experiments. A further and deeper analysis on them serves also for the definition and refinement of CPs and *Heuristics*. Finally, the framework for mapping evolution is formally defined and a software tool implemented for evaluation purposes. In the evaluation method we aim at comparing the adapted mappings, as outcome of applying the proposed framework, with mappings generated by approaches totally based on matching techniques. Different measures will be observed regarding, for instance, the adaptation actions used, the quantity of mapping candidates involved and their semantic correctness. A qualitative evaluation of adapted mappings will also be conducted with experts of the domain.

4 Preliminary Results and Future Work

Empirical basis. We have empirically studied the impact of KOS evolution on mappings by observing the evolution of official mappings between biomedical KOS. We investigate different aspects of changes in KOS elements aiming at understanding the correlations between KOS (complex) changes and how mappings are adapted. Different cases of mapping evolution are considered in the context of KOS complex changes, observing their influence on the changes applied in mappings. Initial results highlight that mappings cannot be adapted according to high level or general types of changes only, but that it is rather necessary to consider fine-grained information on the affecting KOS changes and mappings. Results have also pointed out that it is feasible to have correlations between KOS changes with actions adapting mappings.
The DyKOSMap framework. We have organized the proposed components of the approach into an initial version of the DyKOSMap framework [7]. Figure 1 presents the components and how they are related one to another. The identification of CPs (1) uses two different versions of a same KOS as input, and a set of aspects is designed to describe and recognize CPs. We aim to determine the instances of CPs that took place between two KOS versions. The mapping evolution mechanism must select (2) the appropriate *Mapping Adaptation Action* having as input the current mappings and the set of identified instances of CPs. We perform that supported by the *Heuristics* to know the most adequate actions to apply on impacted mappings. In the last step up-to-date mappings and their history are generated (3) as outcome.
Future work. It involves the refinement of CPs and their identification between KOS versions reusing software tools already available for this purpose. We aim at defining, formalizing and implementing the *Mapping Adaptation Action* and *Heuristics*

computationally. The prototype for mapping evolution shall be developed. Finally, the evaluation will be conducted assessing the results provided by testing the prototype.

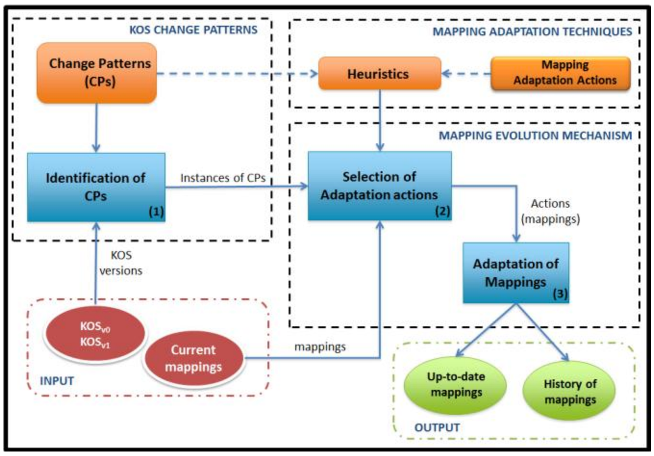

Figure 1. The DyKOSMap framework for supporting mapping evolution

Acknowledgments. FNR of Luxembourg (grant C10/IS/786147 - DynaMO project).

References

1. Hodge, G.: Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files. The Digital Library Federation - Council on Library and Information Resources, Washington, DC (2000).
2. Pavel, S., Euzenat, J.: Ontology Matching: State of the Art and Future Challenges. IEEE Transactions on Knowledge and Data Engineering. 25, 158–176 (2013).
3. Klein, M.: Change Management for Distributed Ontologies, PhD Thesis, Vrije University (2004).
4. Kitamura, Y., Segawa, S.: Deep Semantic Mapping between Functional Taxonomies for Interoperable Semantic Search. In ASWC. pp. 137–151. (2008).
5. Lambrix, P., Strömbäch, L., Tan, H.: Information Integration in Bioinformatics with Ontologies and Standards. In: Bry, F. and Maluszynski, J. (eds.) Semantic Techniques for the Web. pp. 343–376. Springer, Berlin / Heidelberg (2009).
6. Hartung, M., Terwilliger, J., Rahm, E.: Recent advances in schema and ontology evolution. In Schema Matching and Mapping. pp. 149–190. Springer-Verlag (2011).
7. Dos Reis, J.C., Pruski, C., Da Silveira, M., Reynaud-Delaître, C.: Analyzing and Supporting the Mapping Maintenance Problem in Biomedical Knowledge Organization Systems. In SIMI Workshop - 9th Extended Semantic Web Conference. pp. 25–36 (2012).
8. Khattak, A.M., Pervez, Z., Latif, K., Lee, S.: Time Efficient Reconciliation of Mappings in Dynamic Web Ontologies. Knowledge-Based Systems. 35, 369–374 (2012).
9. Velegrakis, Y., Miller, J., Popa, L.: Preserving mapping consistency under schema changes. VLDB Journal. 13, 274–293 (2004).
10. Martins, H., Silva, N.: A User-Driven and a Semantic-Based Ontology Mapping Evolution Approach. 11th ICEIS. pp. 6–10. , Milan (2009).
11. Gross, A., Hartung, M., Thor, A., Rahm, E.: How do computed ontology mappings evolve ? A case study for life science ontologies. Joint Workshop on Knowledge Evolution and Ontology Dynamics (ISWC). pp. 1–12 (2012).